Deep learning to stratify lung nodules on annual follow-up CT

Heuvelmans, Marjolein A.; Oudkerk, Matthijs

Published in:
Lancet digital health

DOI:
10.1016/S2589-7500(19)30156-6

Publication date:
2019

Deep learning to stratify lung nodules on annual follow-up CT

Published Online
October 17, 2019
https://doi.org/10.1016/S2589-7500(19)30156-6

See Articles page e353

For the **DeepLR algorithm** see https://www.caced.jhu.edu

The main goal of lung cancer screening is to identify early-stage lung cancer while preventing unnecessary workup for benign nodules. The major source of error in early lung cancer detection and lung cancer screening is through radiological analysis. Findings of two large randomised controlled trials in high-risk populations (the National Lung Screening Trial [NLST][1] and the NELSON study[2]) have shown the positive effect of lung cancer screening by low-dose chest CT. With expected worldwide implementation of this screening method, the number of CT-detected lung nodules will increase greatly, because around half of people undergoing screening have at least one nodule.[3]

In the USA, nodules detected at screening are managed according to the Lung CT Screening Reporting & Data System (Lung-RADS), which is based on diameter at first detection and the increase in diameter at follow-up. A proposed European protocol[4] is based on nodule volume at first detection and a combination of volume and volume doubling time (VDT) at follow-up. To estimate baseline lung cancer risk, risk calculators are available.[5] However, the probability of nodules detected at baseline being malignant differs from that of nodules detected at incidence screening. Baseline nodules might have been present for years whereas new nodules identified at incidence screening are relatively young and fast growing and possess a substantially higher cancer probability.[6] For new nodules, temporal characteristics (ie, growth or change in density) will provide information about lung cancer risk rather than spatial (size) characteristics. Therefore, new nodules are managed differently from baseline nodules in established guidelines.[1,4] In a lung cancer screening programme, a participant will receive one baseline screening, after which they will have up to 24 annual follow-up CTs. Therefore, accurate management of incident nodules will be crucial for the performance of screening, with more stringent volume threshold criteria.[6]

Machine learning techniques for clinicians' support are augmenting 21st century health care with surprising force.[7] In *The Lancet Digital Health*, Peng Huang and colleagues[8] report a deep learning algorithm (termed DeepLR) for classification of screen-detected lung nodules in follow-up imaging, and they validate DeepLR in a large external dataset. With the DeepLR algorithm, which is available online, Huang and colleagues claim to outperform Lung-RADS and stand-alone diameter-based VDT for lung cancer prediction at follow-up CT, both in a training set (using data from NLST) and in a large external validation set.

In a large-scale contribution to estimate lung cancer risk by deep learning based on images from subsequent CTs,[9] lung cancer risk estimation was restricted to 1-year post CT. Thus, those findings cannot be used to identify individuals who might benefit from a longer screening interval. Huang and colleagues' study adds value[8] because DeepLR can identify a low-risk group among their high-risk screening population who had only a 0·2% chance to develop lung cancer in the next 3 years. These individuals might, therefore, benefit from repeat screening after 2 years, or even 3 years, rather than the current recommendation for 1-year screening. The study findings confirm that their deep learning method, which was trained on time-dependent characteristics, outperforms a diameter-based nodule protocol in terms of lung cancer detection sensitivity.

For **Lung-RADS** see https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads

Before we begin using deep learning techniques as guidance for lung cancer risk estimation in clinical practice, it is important to realise the limitations. First, Huang and colleagues emphasise that DeepLR was trained using mainly baseline and annual screens from the NLST, but the performance of DeepLR on shorter or longer follow-up intervals is unknown. Second, VDT in Huang and colleagues' study was calculated based on manual diameter measurements, which has been shown to be unreliable in a previous study,[10] suggesting that VDT should be based on nodule volume alone. Furthermore, VDT should never be used as a stand-alone procedure, as was done by Huang and colleagues, but always in combination with a nodule volume cutoff.[6] Moreover, training of DeepLR was time-dependent and, therefore, included annual nodule growth rate, which is comparable with a nodule's VDT. Thus, results for VDT in Huang and colleagues' study may not reflect the true VDT value and findings cannot be translated into clinical practice.

What is the main message of Huang and colleagues' study of deep learning in lung cancer nodule stratification? In follow-up imaging of a lung nodule, temporal changes provide valuable additional

information for lung cancer risk prediction to spatial characteristics, surpassing Lung-RADS. The time-dependent training of DeepLR resulted in a very high true-negative nodule rate, potentially identifying individuals who might benefit from repeat screening in 2 or 3 years, compared with the current 1-year recommendation.

Marjolein A Heuvelmans, Matthijs Oudkerk
University Medical Center Groningen, Department of Epidemiology (MAH), and Faculty of Medical Sciences (MO), University of Groningen, 9700 RB Groningen, Netherlands; Department of Pulmonology, Medisch Spectrum Twente, Enschede, Netherlands (MAH); and Institute for DiagNostic Accuracy, Groningen, Netherlands (MO)
m.a.heuvelmans@umcg.nl

MAH and MO declare no competing interests.

1 National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; **365:** 395–409.
2 De Koning H, Van Der Aalst C, Ten Haaf K, Oudkerk M. PL02.05 Effects of volume CT lung cancer screening: mortality results of the NELSON randomised-controlled population based trial. *J Thorac Oncol* 2018; **13** (suppl)**:** S185.
3 Horeweg N, van Rosmalen J, Heuvelmans MA, et al. Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. *Lancet Oncol* 2014; **15:** 1332–41.
4 Oudkerk M, Devaraj A, Vliegenthart R, et al. European position statement on lung cancer screening. *Lancet Oncol* 2017; **18:** e754–66.
5 McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 2013; **369:** 910–19.
6 Walter JE, Heuvelmans MA, Ten Haaf K, et al. Persisting new nodules in incidence rounds of the NELSON CT lung cancer screening study. *Thorax* 2019; **74:** 247–53.
7 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019; **25:** 44–56.
8 Huang P, Lin CT, Li Y, et al. Prediction of lung cancer risk at follow-up screening with low-dose CT: a training and validation study of a deep learning method. *Lancet Digital Health* 2019; published online Oct 17. https://doi.org/10.1016/S2589-7500(19)30159-1.
9 Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019; **25:** 954–61.
10 Heuvelmans MA, Walter JE, Vliegenthart R, et al. Disagreement of diameter and volume measurements for pulmonary nodule size estimation in CT lung cancer screening. *Thorax* 2018; **73:** 779–81.